The Stationary Phase Approximation, Time-Frequency Decomposition and Auditory Processing

The Stationary Phase Approximation, Time-Frequency Decomposition and Auditory Processing

Bernard Mulgrew, *Fellow, IEEE*

***Abstract*—The principle of stationary phase (PSP) is re-examined in the context of linear time-frequency (TF) decomposition using Gaussian, gammatone and gamma chirp filters at uniform, logarithmic and cochlear spacings in frequency. This necessitates consideration of the use the PSP on non-asymptotic integrals and leads to the introduction of a test for phase rate dominance. Regions of the TF plane that pass the test and do not contain stationary phase points contribute little or nothing to the final output. Analysis values that lie in these regions can thus be set to zero, i.e., sparsity. In regions of the TF plane that fail the test or are in the vicinity of stationary phase points, synthesis is performed in the usual way. A new interpretation of the location parameters associated with the synthesis filters leads to: i) a new method for locating stationary phase points in the TF plane and ii) a test for phase rate dominance in that plane. Together this is a TF stationary phase approximation (TFSPA) for both analysis and synthesis. The stationary phase regions of several elementary signals are identified theoretically and examples of reconstruction given. An analysis of the TF phase rate characteristics for the case of two simultaneous tones predicts and quantifies a form of simultaneous masking similar to that which characterizes the auditory system.**

***Index Terms*—Method of reassignment, cochlear filters, gammatone, gammachirp, simultaneous masking.**

I. Introduction

THE principle (or method) of stationary phase (PSP) [1] is a result from asymptotics that can provide closed-form approximations, in the limit as $\lambda \to \infty$, to often intractable oscillatory integrals of the form

$$I = \int_{-\infty}^{\infty} a(t) \mathrm{e}^{\jmath \lambda b(t)} \mathrm{d}t \tag{1}$$

where $a, b, t, \lambda \in \mathbb{R}$. There are two steps involved: PSP(i) recognition that in the limit the integral will be almost zero everywhere in the interval $\{t : -\infty \leq t \leq \infty\}$ except near values of t where the derivative $\dot{b}(t)$ is zero, the stationary phase points; PSP(ii) the integrand in the vicinity of these stationary phase points can be expressed in terms of the second derivative of the phase i.e., $\ddot{b}(t)$. Perhaps the most successful application of the PSP in signal processing has been in the context of synthetic aperture radar (SAR), where it is the starting point in the development of many of the Fourier-based imaging algorithms, cf. [2]. Application of the PSP is not without its pitfalls. It is tempting to use the PSP in the non-asymptotic cases where $\lambda = 1$ to find closed form approximations to integrals such as Fourier transforms. The argument for this requires that the phase $b(t)$ is changing much more rapidly than the amplitude $a(t)$. However, as pointed out in [3], a degree of care must be exercised, particularly with PSP(ii).

The primary interest here is its application to linear time-frequency (TF) decomposition [4]. The motivation is the recent resurgence of interest in analogue filter banks both as part of a synthetic cochlea and as a means to provide power efficient implementations of analysis filter banks [5]. The desire with both is to extract salient features from the TF decomposition using the limited functionality associated with analogue circuitry. This does not deny the considerable work that has been done on computational modelling of the auditory system, typified by papers such as [6] and the references therein. However the main purpose of such work is to model and predict the response of the auditory system to stimulus rather than to expose the signal processing principles that might be at work.

The PSP is a natural place to start because of the prevalence of oscillatory terms in TF decompositions and the hope that the stationary phase points may provide a means for identifying salient features as well as a focus for sparse decompositions without the need for the usual iterative re-synthesize steps, cf. [4] chapter 12. The PSP has been applied to linear TF decomposition for both analysis, [7] and [8], and synthesis [9], the latter leading to the method of reassignment. Subsequent developments of the method of reassignment are documented in [10]. Reassignment can provide improved estimates of the location of components in the TF plane. Related methods such as [11] and [12] restrict this improvement to the frequency direction alone. In particular synchrosqueezing [13] can be viewed as a form of reassignment that also facilitates sparse re-synthesis.

The objective here is to provide similar capabilities, i.e., location accuracy and sparsity of representation, using the limited functionality of analogue filter banks and, by doing so, to shed some light on the principles that might be at play within the cochlea which has similar limited functionality. The approach adopted is to revisit [9] and to fundamentally re-interpret it to provide a PSP-based approximation to the TF synthesis integral. There is no attempt to either reassign [9] or relocate [13] components in the TF plane because of the said limited functionality.

Manuscript received June 24, 2013; accepted September 09, 2013. Date of publication October 04, 2013; date of current version December 03, 2013. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Lawrence Carin. This work was supported by the Royal Academy of Engineering, Selex ES and EPSRC (EP/J015180/1).

The author is with the Institute for Digital Communications, School of Engineering, The University of Edinburgh, Edinburgh, EH9 3JL, U.K. (e-mail: B.Mulgrew@ed.ac.uk).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TSP.2013.2284479

Another alternative would be to follow an amplitude-based approach such as ridgelets [14]. However this leads to algorithms that are far from feasible with analogue circuitry and further, ridges in the TF plane may not be appropriate when dealing with auditory filters such as the gammatone [15] and gammachirp [16] which have asymmetrical impulse responses and, in the case of the latter, an asymmetrical frequency response. Concentration on the synthesis rather than the analysis integral is advocated because: (i) most methods for sparse atomic decomposition, cf. [4] chapter 12, are based on re-synthesis of the original waveform; (ii) one of the main functions of the auditory system is to code the incident waveforms and coding requires at least some consideration of the potential for reconstruction (even when reconstruction is not a requirement); (iii) smooth variations of the magnitude and phase of the integrand are more readily satisfied for the synthesis integral than the analysis integral (because the signal has already been filtered in the analysis process).

The primary contribution of this paper is a TF stationary phase approximation (TFSPA) for both linear signal analysis and synthesis (Section IV). It reduces both the extraction of salient features from TF analysis and the selection of components for synthesis to simple tests that can be performed instantaneously at the output of an analysis filter bank. Supporting contributions include:

- A re-examination of the application of the PSP to non-asymptotic integrals (Section III) that leads to the use of PSP(i) without the need for PSP(ii). The subsequent introduction of the concept of phase rate dominance that partitions the interval into sets where the PSP can and cannot be applied.
- A new interpretation of the time-location parameter associated with the synthesis double-integral that leads to: (i) the development of both a simple and a constant false alarm rate (CFAR) test for stationary phase points in the TF plane (Section IV); (ii) the extension of the phase-rate dominance concepts from one to 2 dimensions (Appendix A) to provide a test for such dominance in the TF plane (Section IV).

In addition, the stationary phase points and phase-rate dominance characteristics of elementary signals such as impulses, phasors, chirps and decaying phasors are analyzed (Section V). This analysis also serves to illuminate the results when TFSPA is applied to a speech signal. Gammatone/chirp filters at cochlear spacing [17] are used throughout to provide insight into auditory processing (Section II). An example of this is the final contribution of the paper:

- An analysis of the performance of the simple stationary-phase-point detector proposed here when two tones are applied simultaneously. This analysis predicts a form of simultaneous masking [18] similar to that which characterizes the auditory system (Section VI).

Aspects of this work were reported briefly in [19].

II. Preliminaries

Consider a linear TF analysis $X(t,\omega)$ of a signal of interest $x(t)$ of the form:

$$X(t,\omega) = x(t) * h_\omega(t), \{\omega : \omega_{\min} < \omega < \omega_{\max}\} \qquad (2)$$

where * denotes convolution and the impulse response of a single filter in the analysis filter bank is given by:

$$h_\omega(t) = \beta h(\beta t)\mathrm{e}^{\jmath\omega t} \qquad (3)$$

Each filter is formed using a prototype filter $h(t)$, a nominal bandwidth β and a frequency location ω. The frequency response $H_\omega(\Omega)$ of the analysis filter is related to the frequency response of the prototype, i.e., $H_\omega(\Omega) = \mathsf{F}[h_\omega(t)] = H\left(\frac{\Omega-\omega}{\beta}\right)$, where $H(\Omega) = \mathsf{F}[h(t)] = \int_{-\infty}^{\infty} h(t)\mathrm{e}^{-\jmath\Omega t}\,\mathrm{d}t$. When dealing with these prototype filters the following notation is used to indicate the derivatives of their impulse and frequency responses, i.e., $\dot{h}(t) = \frac{\mathrm{d}}{\mathrm{d}t}h(t)$ and $\dot{H}(\Omega) = \frac{\mathrm{d}}{\mathrm{d}\Omega}H(\Omega)$ respectively. Gaussian, gammatone and gammachirp prototypes are considered here. While a Gaussian prototype is a common [4] and analytically convenient choice for TF analysis, the gammatone [15] and gammachirp prototypes [16] more closely model the cochlea in the ear. The prototype filters are normalized such that $H(0) = 1$. This property and multiplication by the nominal bandwidth β in (3) normalizes the maximum gain of each filter to unity at ω rad/s. A re-synthesis, $\hat{x}(t)$, of the signal of interest is performed using filters matched to $h_\omega(t)$. Thus for real signals of interest

$$\hat{x}(t) = \frac{1}{C}\Re\left\{\int_0^1 \int_{-\infty}^{\infty} Z(\tau,\mu)\mathrm{d}\tau\mathrm{d}\mu\right\} \qquad (4)$$

with integrand $Z(\tau,\mu) = X(\tau,\omega)h_\omega^*(\tau - t)$ and where C is a constant and $\Re\{\cdot\}$ and $\Im\{\cdot\}$ denote 'real part of' and 'imaginary part of' respectively. The time variable t is suppressed in this definition of the integrand to emphasize that the integration is respect to τ and the filter bank variable μ. The latter lies in the range $[0,1]$ and is a monotonic function of ω. It provides a convenient way of dealing with a number of possible filter bank spacings. A value $\mu = 0$ indicates the lower edge of the filter bank and $\mu = 1$ indicates the upper edge. The frequency ω at which the filters gain is a maximum is a function of μ and the bandwidth of the filter β is proportional to the derivative $\frac{\mathrm{d}\omega}{\mathrm{d}\mu}$, i.e., $\beta \propto \frac{\mathrm{d}\omega}{\mathrm{d}\mu}$. Hence the total number of filters required to just cover the band of interest is given approximately by $\frac{\mathrm{d}\omega}{\mathrm{d}\mu}\frac{1}{\beta}$. In the following $\omega_{\min}$ is nominally the lowest frequency covered by the filter bank and $\omega_{\max}$ is the maximum. For a uniformly spaced filter bank $\omega = \{\omega_{\max} - \omega_{\min}\}\mu + \omega_{\min}$ and hence $\mu = \{\omega - \omega_{\min}\}/\{\omega_{\max} - \omega_{\min}\}$, $\frac{\mathrm{d}\omega}{\mathrm{d}\mu} = \omega_{\max} - \omega_{\min}$, the nominal bandwidth β is a constant and $\frac{\mathrm{d}\beta}{\mathrm{d}\omega} = 0$. For logarithmically spaced filter banks, similar to wavelets [4], $\omega = \omega_{\min}\mathrm{e}^{b\mu}$ where $b = \ln\left(\frac{\omega_{\max}}{\omega_{\min}}\right)$ and hence $\mu = \frac{1}{b}\ln\left(\frac{\omega}{\omega_{\min}}\right)$, $\frac{\mathrm{d}\omega}{\mathrm{d}\mu} = b\omega$, the nominal bandwidth β is proportional to ω and $\frac{\mathrm{d}\beta}{\mathrm{d}\omega}$ is a constant. For a cochlear spaced filter banks, based on the approximation of [17] for low sound pressure levels, $\omega = \{\omega_{\min} + a\}\mathrm{e}^{b\mu} - a$, where $a = \frac{2\pi\times 10^3}{4.37}$, $b = \ln(\frac{\omega_{\max}+a}{\omega_{\min}+a})$ and hence $\mu = \frac{1}{b}\ln\left(\frac{\omega+a}{\omega_{\min}+a}\right)$, $\frac{\mathrm{d}\omega}{\mathrm{d}\mu} = b\{\omega + a\}$ and $\frac{\mathrm{d}\beta}{\mathrm{d}\omega}$ is a constant. For all the above filter banks, the equivalent rectangular bandwidth (ERB) [18] and the 3 dB bandwidth are related to the nominal bandwidth β in a straightforward way.

As indicated, the Gaussian, gammatone and gammachirp prototype filters all have a maximum gain of unity at zero frequency. They are scaled in time and modulated in frequency

to form the various filter banks. A *Gaussian prototype* has a non-causal impulse response $h(t) = \frac{1}{\sqrt{2\pi}}\mathrm{e}^{-t^2/2}$ with a peak at $t = 0$, a group delay of 0 and a derivative $\dot{h}(t) = -th(t)$. Its frequency response is $H(\Omega) = \mathrm{e}^{-\frac{\Omega^2}{2}}$ with derivative $\dot{H}(\Omega) = -\Omega H(\Omega)$. An order n *gammachirp prototype*, with chirp rate parameter c, has a causal impulse response defined in terms of the complex gamma function $\Gamma(\,\cdot\,)$ as

$$h(t) \triangleq \frac{\left\{1 + \jmath\frac{c}{n}\right\}^{n+\jmath c}}{\Gamma(n+\jmath c)} t^{n-1}\mathrm{e}^{-\{1+\jmath\frac{c}{n}\}t+\jmath c\ln(t)},\ \forall t \geq 0$$

and $h(t) = 0$ otherwise. This is a minor modification to the gammachirp filter of [16] in order to decouple the dependency between the location of the peak gain in the frequency response and the chirp rate parameter c. The time derivative is

$$\dot{h}(t) = \left\{-1 + \frac{n-1}{t} + \jmath c\left\{\frac{1}{t} - \frac{1}{n}\right\}\right\} h(t). \tag{5}$$

The gammatone prototype is a specific case of this when the chirp rate parameter, c, is zero. The gammachirp prototype has frequency response

$$H(\Omega) = \frac{\left\{\jmath\frac{c}{n} + 1\right\}^{n+\jmath c}}{\left\{\jmath\left\{\Omega + \frac{c}{n}\right\} + 1\right\}^{n+\jmath c}},$$

the derivative of which is

$$\dot{H}(\Omega) = \frac{\{c - \jmath n\}}{\left\{\jmath\left\{\Omega + \frac{c}{n}\right\} + 1\right\}} H(\Omega).$$

The peak in the magnitude of the impulse response occurs at $t = n - 1$ while the group delay of the filter at $\Omega = 0$ is $-\frac{\mathrm{d}\angle H(0)}{\mathrm{d}\Omega} = n$, where $\angle H$ indicates the phase of H. The gammachirp prototype, like the Gaussian, is scaled (in time) and modulated to give the analysis filter of (3). The chirp rate parameter is unaffected by the time scaling because of the action of the logarithm function i.e., $\jmath c\ln(\beta t) = \jmath c\ln(\beta) + \jmath c\ln(t)$. The time scaling adds a phase shift $c\ln(\beta)$ to the impulse response. Thus values of c can be used interchangeably with [16]. From (5), the phase derivative of the impulse response is: $\frac{\mathrm{d}\angle h(t)}{\mathrm{d}t} = -\frac{c}{n} + \frac{c}{t}$, which is zero at the group delay of $t = n$. Likewise $\dot{h}(t) = -\frac{1}{n}$ at the group delay.

III. The PSP and Non-Asymptotic Integrals

Consider the integral (1) evaluated over an interval $[t - \Delta/2, t + \Delta/2]$. Assuming that the function $f(t) = a(t)\mathrm{e}^{\jmath\lambda b(t)}$ is well approximated by its Taylor series over this interval, gives

$$\begin{aligned} I_\Delta(t) &\triangleq \int_{t-\Delta/2}^{t+\Delta/2} f(t')\mathrm{d}t' \\ &\approx f(t)\int_{t-\Delta/2}^{t+\Delta/2}\left\{1 + \frac{\dot{f}(t)}{f(t)}\{t'-t\} + \frac{\ddot{f}(t)}{f(t)}\frac{\{t'-t\}^2}{2}\right\}\mathrm{d}t' \end{aligned}$$

The derivatives can expressed as

$$\frac{\dot{f}(t)}{f(t)} = \left\{\frac{\dot{a}(t)}{a(t)} + \jmath\lambda\dot{b}(t)\right\}$$

and

$$\frac{\ddot{f}(t)}{f(t)} = \frac{\ddot{a}(t)}{a(t)} - \left\{\frac{\dot{a}(t)}{a(t)}\right\}^2 + \jmath\lambda\ddot{b}(t) + \left\{\frac{\dot{a}(t)}{a(t)} + \jmath\lambda\dot{b}(t)\right\}^2$$

In the asymptotic case, at a suitably large value of λ, $\frac{\dot{f}(t)}{f(t)} \approx \jmath\lambda\dot{b}(t)$ and $\frac{\ddot{f}(t)}{f(t)} \approx -\lambda^2\dot{b}^2(t) + \jmath\lambda\ddot{b}(t)$ and hence the integral is dominated by the phase derivatives $\dot{b}(t)$ and $\ddot{b}(t)$. Thus, at a stationary point, where $\dot{b}(t) = 0$, the integral can be evaluated in terms of $\ddot{b}(t)$ without reference to derivatives of $a(t)$. For the non-asymptotic case where $\lambda = 1$, these approximations are also dependent on the relationships between the derivatives of the magnitude and phase of the integrand. Thus $\frac{\dot{f}(t)}{f(t)} \approx \jmath\dot{b}(t)$ provided, as in [7], that

$$|\,\dot{b}(t)\,| \gg \left|\frac{\dot{a}(t)}{a(t)}\right| \tag{6}$$

and $\frac{\ddot{f}(t)}{f(t)} \approx -\dot{b}^2(t) + \jmath\ddot{b}(t)$ provided *also* that

$$|\,\ddot{b}(t)\,| \gg \left|\frac{\ddot{a}(t)}{a(t)} - \left\{\frac{\dot{a}(t)}{a(t)}\right\}^2\right| = \left|\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\dot{a}(t)}{a(t)}\right)\right| \tag{7}$$

If the amplitude and phase derivatives are available then at every value of t it is possible to test for what might be called *first order or second order phase-rate dominance* using (6) or, (6) and (7), respectively. The approach adopted here, where a closed form approximation to the integral is not required, is to circumvent the difficulties associated with PSP(ii) by simply not using that step.

Let the set of stationary phase points be denoted by $\{t_i\}_{i\in\{i:\dot{b}(t_i)=0\}}$ (or more compactly $\{t_i\}_i$) and let the interval that contains the ith stationary phase point be $S_i \in \{t : t_i - \delta_1 < t < t_i + \delta_2, \delta_1 \geq 0, \delta_2 \geq 0\}$. Further, the inequality (6) defines a set S_0 that is the union of intervals of t where PSP(i) is valid and the compliment to that set $\bar{S}_0$ where it is not. The integral over the whole real line, (1), can then be replaced by an integral over the union of intervals such as S_i and $\bar{S}_0$, i.e., an integral over $S \in \{\bigcup_i S_i\} \cup \bar{S}_0$. It is also worth noting that for (6) to be satisfied at or near a stationary phase point, would also require that $|\frac{\dot{a}(t)}{a(t)}| \to 0$ as $t \to t_i$. Thus the normalized amplitude rate must go to zero more rapidly than the phase rate, i.e., (7) must apply. Thus there are liable to be intervals where $S_i \in \bar{S}_0$ and for these intervals there is no need to identify δ_1 and δ_2.

IV. A Time-Frequency Stationary Phase Approximation

The PSP can be extended to double integrals such as (4) with the result defined in terms of the gradient and Hessian of the phase of the integrand, cf. [1], pg. 478. However, as in Section III the full form of the PSP is not used here. Specifically, the gradient is used in PSP(i) to identify the stationary phase points but the Hessian of PSP(ii) is not used to form an approximation to the integral in the vicinity of these points. A test similar to (6) is developed for double integrals in Appendix A to identify regions of phase-rate dominance in the TF plane where the PSP is applied by numerical integration in the vicinity of the stationary phase points. As in Section III, it is convenient

to define normalized time- and frequency- derivatives of the integrand $Z(\tau,\mu)$, i.e., $Z_\tau \triangleq \frac{\frac{\partial}{\partial\tau}Z(\tau,\mu)}{Z(\tau,\mu)}$ and $Z_\mu \triangleq \frac{\frac{\partial}{\partial\omega}Z(\tau,\mu)\frac{\mathrm{d}\omega}{\mathrm{d}\mu}}{Z(\tau,\mu)}$ respectively, where

$$\begin{aligned} Z_\tau &= \frac{\frac{\partial}{\partial\tau}X(\tau,\omega)}{X(\tau,\omega)} + \frac{\frac{\partial}{\partial\tau}h^*_\omega(\tau-t)}{h^*_\omega(\tau-t)} \\ &= \frac{\frac{\partial}{\partial\tau}X(\tau,\omega)}{X(\tau,\omega)} + \beta\frac{\dot{h}^*(\beta\{\tau-t\})}{h^*(\beta\{\tau-t\})} - \jmath\omega \end{aligned} \tag{8}$$

and

$$Z_\mu = \left\{\frac{\frac{\partial}{\partial\omega}X(\tau,\omega)}{X(\tau,\omega)} + \frac{\frac{\partial}{\partial\omega}h^*_\omega(\tau-t)}{h^*_\omega(\tau-t)}\right\}\frac{\mathrm{d}\omega}{\mathrm{d}\mu} \tag{9}$$

where

$$\begin{aligned} &\frac{\frac{\partial}{\partial\omega}h^*_\omega(\tau-t)}{h^*_\omega(\tau-t)}l \\ &\quad= \frac{\mathrm{d}\beta}{\mathrm{d}\omega}\left\{\frac{1}{\beta} + \frac{\dot{h}^*(\beta\{\tau-t\})}{h^*(\beta\{\tau-t\})}\{\tau-t\}\right\} - \jmath\{\tau-t\} \end{aligned} \tag{10}$$

Both Z_τ and Z_μ are additions of a signal dependent term, e.g., $\frac{\frac{\partial}{\partial\tau}X(\tau,\omega)}{X(\tau,\omega)}$, and a signal independent term, e.g., $\frac{\frac{\partial}{\partial\tau}h^*_\omega(\tau-t)}{h^*_\omega(\tau-t)}$. The former is a function of the pair (τ,ω) whereas the latter is a function of the pair $(\tau-t,\omega)$. Note that ω and $\{\tau-t\}$ are themselves parameters of the filter, specifically, the frequency where the filter has maximum gain ω and the delay $\{\tau-t\}$ between the input and output of the filter. In contrast to the method of re-assignment derived in [9], the delay term $\{\tau-t\}$ is interpreted here as the group delay

$$\{\tau-t\} = g(\omega) \triangleq -\frac{\mathrm{d}}{\mathrm{d}\Omega}\angle H_\omega(\Omega)|_{\Omega=\omega} \tag{11}$$

of the filter (3) at frequency ω. Simple expressions for the group delay are given in Section II in terms of the group delays of the prototype filters at zero frequency. The justification for the use of (11) in (8), (9) and (10) proceeds as follows: $\{\tau-t\}$ is the delay between the input signal $x(t)$ and the output of the analysis filter $X(\tau,\omega)$; since each filter is tuned to have a maximum gain at a particular frequency ω, the delay through the filter is also tuned to the rate of change of the phase response at the frequency where the gain is maximum. A particular filter is thus jointly labeled with both its frequency ω and the group delay at that frequency $g(\omega)$. Thus (8) and (9) can be written as

$$Z_\tau = \frac{\frac{\partial}{\partial\tau}X(\tau,\omega)}{X(\tau,\omega)} + \beta\eta(\omega) - \jmath\omega \tag{12}$$

and

$$Z_\mu = \left\{\frac{\frac{\partial}{\partial\omega}X(\tau,\omega)}{X(\tau,\omega)} + \frac{\mathrm{d}\beta}{\mathrm{d}\omega}\left\{\frac{1}{\beta} + \eta(\omega)g(\omega)\right\} - \jmath g(\omega)\right\}\frac{\mathrm{d}\omega}{\mathrm{d}\mu} \tag{13}$$

respectively, where $\eta(\omega) = \frac{\dot{h}^*(\beta\{g(\omega)\})}{h^*(\beta\{g(\omega)\})}$. Then because $\Im\{\eta(\omega)\} = 0$ for all three filter types *including the complex gammachirp* (cf. Section II), the time derivative of the phase of the integrand is

$$\Im\{Z_\tau\} = \Im\left\{\frac{\frac{\partial}{\partial\tau}X(\tau,\omega)}{X(\tau,\omega)}\right\} - \omega \tag{14}$$

and the frequency derivative is

$$\Im\{Z_\mu\} = \left\{\Im\left\{\frac{\frac{\partial}{\partial\omega}X(\tau,\omega)}{X(\tau,\omega)}\right\} - g(\omega)\right\}\frac{\mathrm{d}\omega}{\mathrm{d}\mu} \tag{15}$$

Time and frequency derivatives of the analysis integral (2) are constructed using the derivative filters $\frac{\partial}{\partial\tau}h_\omega(\tau)$ and $\frac{\partial}{\partial\omega}h_\omega(\tau)$ respectively, cf. [20], [9]:

$$\frac{\partial}{\partial\tau}X(\tau,\omega) = x(\tau) * \frac{\partial}{\partial\tau}h_\omega(\tau) \tag{16}$$

$$\frac{\partial}{\partial\omega}X(\tau,\omega) = x(\tau) * \frac{\partial}{\partial\omega}h_\omega(\tau) \tag{17}$$

Alternatively, as suggested in [9], they can be derived from the output of (2) by direct differentiation of $X(\tau,\omega)$ to obtain (16) and by using neighboring analysis filters to obtain an approximation to (17). Stationary phase points $\{(\tau_i,\omega_i)\}_i$ are solutions to:

$$\Im\{Z_\tau\} = \Im\{Z_\mu\} = 0 \tag{18}$$

The derivative filters can be expressed in terms of the prototype filter $h(t)$ and its derivative $\dot{h}(t)$. Expressions for the derivative $\dot{h}(t)$ of the Gaussian and gammachirp filters can be found in Section II. From (18) there are two conditions that must be satisfied simultaneously for a stationary phase point to occur at (ω_i,τ_i), specifically:

1) the frequency, $\frac{\partial\angle X(\tau,\omega)}{\partial\tau} = \Im\left\{\frac{\frac{\partial}{\partial\tau}X(\tau,\omega)}{X(\tau,\omega)}\right\}$, observed at the output of filter at time τ, is equal to the centre frequency ω of the filter;
2) the delay, $\frac{\partial\angle X(\tau,\omega)}{\partial\omega} = \Im\left\{\frac{\frac{\partial}{\partial\omega}X(\tau,\omega)}{X(\tau,\omega)}\right\}$, observed at the output of the filter at frequency ω, is equal to the group delay $g(\omega)$ of the filter at that frequency.

Together these define a *signal matching condition*: at (τ,ω) the frequency and delay observed at the output of the filter must match the designed centre frequency and group delay of that filter.

Locating stationary phase point requires a grid search over ω for a bank of analogue filters or over both ω and τ, for a discrete-time filter bank. Such a grid search is not onerous since it is implicit in the implementation of the analysis integral. With a grid search there is always the risk of missing the pair (τ_i,μ_i) that satisfy (18). This risk can be reduced by: (i) defining a phase gradient vector

$$\boldsymbol{\phi}(\tau,\mu) \triangleq [\Im\{Z_\tau\}\ \Im\{Z_\mu\}]^T \tag{19}$$

where the superscript T indicates matrix transpose; (ii) using the Euclidean norm of this vector to construct a test for stationary phase points, i.e.,

$$\|\boldsymbol{\phi}(\tau,\mu)\| < C_1 \tag{20}$$

where the threshold C_1 is a small positive real constant. The Euclidean norm is used here for analytic convenience when dealing with deterministic signals. Other vector norms may be appropriate and may have desirable properties with respect to ease of implementation.

In noisy environments any detector will make type I errors (false alarms) and it is common to design the detector to operate at a specified false alarm rate [21]. In [19], closed form

expressions for the probability density function (PDF) and cumulative distribution function (CDF) of $\Im\{Z_\tau\}$ and $\Im\{Z_\mu\}$ are developed under the null hypothesis that $x(t)$ is Gaussian white noise. Based on this a CFAR detector for stationary phase points is proposed here using the test:

$$|\Im\{Z_\tau\}| < T_\tau(\mu) \wedge |\Im\{Z_\mu\}| < T_\mu(\mu) \tag{21}$$

where $\wedge$ indicates the logical "AND" operation. Details of the evaluation of the thresholds $T_\tau(\mu)$ and $T_\mu(\mu)$ together with the assumptions used in their derivation are provided in Appendix B. For a non-uniform filterbank the values of the thresholds required to achieve a particular false alarm rate are dependent on the frequency location of the filter and thus these thresholds are denoted as functions of μ.

In addition to finding stationary phase points, (12) and (13) can also be used to test for phase-rate dominance in the TF plane. For phase-rate dominance the inequality

$$\frac{p(\tau,\mu)}{\|\mathbf{a}(\tau,\mu)\|} > C_2 \tag{22}$$

must be satisfied, where the amplitude gradient vector is

$$\mathbf{a}(\tau,\mu) \triangleq [\Re\{Z_\tau\}\ \Re\{Z_\mu\}]^T \tag{23}$$

and the threshold C_2 is a positive real constant greater than or equal to one. The projection term

$$p(\tau,\mu) = \frac{|\boldsymbol{\phi}^T(\tau,\mu)\mathbf{W}\mathbf{a}(\tau,\mu)|}{\|\mathbf{a}(\tau,\mu)\|} \tag{24}$$

is formed from the sum of the projection of the phase rate vector in the direction of the normalized amplitude rate, i.e., $\boldsymbol{\phi}^T\mathbf{a}/\|\mathbf{a}\|$, plus the projection in the orthogonal direction, i.e., $\boldsymbol{\phi}^T\left[\begin{smallmatrix}0 & 1\\ -1 & 0\end{smallmatrix}\right]\mathbf{a}/\|\mathbf{a}\|$ and hence $\mathbf{W} = \left[\begin{smallmatrix}1 & 1\\ -1 & 1\end{smallmatrix}\right]$. The test is derived in Appendix A. Thus the PSP divides the TF plane into two regions: a region S_0 where (22) is satisfied and the rest of the TF plane $\tilde{S}_0$ where it is not.

Given the stationary phase points $\{(\tau_i,\mu_i)\}_i$ that are solutions to (18), the stationary phase approximation is invoked by replacing (4) by:

$$\hat{x}(t) \approx \frac{1}{C}\Re\left\{\iint_S X(\tau,\omega)h^*_\omega(\tau-t)\,\mathrm{d}\mu\mathrm{d}\tau\right\} \tag{25}$$

where S is a subset of the TF plane defined as $S = \{\bigcup_i S_i\}\cup\bar{S}_0$, S_i is the neighborhood of the ith stationary phase point such that $(\mu_i,\tau_i) \in S_i$ and $\bigcup_i S_i$ contains all points in the TF plane that satisfy (20) (or (21)). Equation (25) promises sparsity directly from analysis without the computationally expensive re-synthesis step associated with most methods for sparse atomic decomposition. The atomic decomposition of (25) is sparse in the sense that $\forall\,(\tau,\mu) \in \bar{S}$, the coefficient $X(\tau,\omega)$ is implicitly set to zero. However there are no guarantees about the degree of sparsity that can be achieved or the quality of reconstruction that might be expected apart from the usual ones that might be expected from a well-designed snug or tight frame [22]. This will be explored in the following section. Together the analysis steps of (2), (12) and (13), the selection inequalities (20) or (21) and (22) and the synthesis equation (25) form what might be called a time-frequency stationary phase approximation (TFSPA).

V. Elementary Signals

This Section is devoted to a consideration of the stationary phase regions of the TF plane associated with some important elementary signals and also in identifying regions of phase-rate dominance. Because of the elementary nature of the signals there is some hope that closed form solutions are possible. These elementary waveforms are: an impulse; a single tone; a linear chirp; a decaying phasor. Closed form solutions that define the stationary phase points are derived. Within the constraints of the available space results of the numerical evaluation of these regions are also provided to confirm and expand upon the theoretical results. Unless otherwise stated, the results relate to order $n = 4$ gammatone ($c = 0$) and gammachirp ($c = 4$) filter banks at the cochlear spacing described in Section II. Numerical evaluation considers a frequency band from $\omega_{\min} = 2\pi \times 100$ rad/s to $\omega_{\max} = 2\pi \times 5000$ rad/s at a sampling rate of 20 kHz. For the gammatone, $\frac{\mathrm{d}\omega}{\mathrm{d}\mu}\frac{1}{\beta} = 25$ and $\frac{\mathrm{d}\beta}{\mathrm{d}\omega} = \frac{1}{9}$ and for the gammachirp, $\frac{\mathrm{d}\omega}{\mathrm{d}\mu}\frac{1}{\beta} = 36$ and $\frac{\mathrm{d}\beta}{\mathrm{d}\omega} = \frac{1}{13}$. These values are commensurate with ERB figures for the cochlea, cf. [17]. The spacing of the filters in frequency is such that there are 4 filters with their maximum gain within the ERB of each filter [22]. In this case 103 filters are used to cover the stated band. The combined frequency response of the analysis and synthesis filters banks has a linear phase and is flat to within a fraction of a dB over the band. Thus, over the band of interest, the combination of analysis and synthesis filter banks, (2) and (4), is effectively distortionless. All filters are approximated by simply truncating them at a suitable point to give finite impulse response (FIR) filters. Anticausal filters (e.g., synthesis filters) are simulated by incorporating suitable delays. For the purposes of display and comparison, the scaling factor C in (25) is evaluated using a least squares fit to the signal.

For the inequalities ((20) and (22)), $C_1 = 10$ and $C_2 = 1$. These values are used in all simulation results presented in this section. The former was found experimentally to provide a good indication of the stationary phase regions for the filter bank described above and for all combinations of filter spacing and prototype filter described in Section II. The latter is an extreme value. The inequalities, (22) of Section III and (48) of Appendix A, suggest a larger value such as $C_2 = 10$. However a value of unity was chosen because: (i) it makes clear the boundary in the TF plane between regions where the projection $p(\tau,\mu)$ is greater than the norm $\|\mathbf{a}(\tau,\mu)\|$; (ii) it reflects a desire to test the degree of sparsity that could be obtained. Thus (22) becomes

$$p(\tau,\mu) > \|\mathbf{a}(\tau,\mu)\| \tag{26}$$

Studies of ridgelets and skeletons as in [14] suggest that for asymptotic signals, very sparse representations are possible.

A. An Impulse

The impulse is obviously important because it is an extreme example of a transient signal. Further it is not an asymptotic signal in the sense considered in [7] and as such illustrates the advantages of applying to PSP to the synthesis integral rather than the analysis integral. For an impulse $x(t) = \delta(t)$ the analysis of (2) yields $X(\tau,\omega) = \beta h(\beta\tau)\mathrm{e}^{\jmath\omega\tau}$. From which, (12) and (13) give

$$Z_\tau = \beta\left\{\frac{\dot{h}(\beta\tau)}{h(\beta\tau)} + \eta\right\} \tag{27}$$

and

$$Z_\mu = \frac{\mathrm{d}\omega}{\mathrm{d}\mu}\frac{1}{\beta}\left\{\frac{\mathrm{d}\beta}{\mathrm{d}\omega}\left\{2+\beta\tau\frac{\dot{h}(\beta\tau)}{h(\beta\tau)} + \beta g(\omega)\eta\right\} + \jmath\{\beta\tau - \beta g(\omega)\}\right\} \tag{28}$$

For a Gaussian prototype $g(\omega) = 0$, $\eta = 0$ and $\frac{\dot{h}(\beta\tau)}{h(\beta\tau)} = -\beta\tau$. For a gammachirp prototype $g(\omega) = \frac{n}{\beta}$, $\eta = -1/n$ and $\frac{\dot{h}(\beta\tau)}{h(\beta\tau)} = -1 + \frac{n-1}{\beta\tau} + \jmath c\{\frac{1}{\beta\tau} - \frac{1}{n}\}$, cf. Section II. Thus, for all three filter types, at the filter spacing considered, the stationary phase region is defined as $\tau = g(\omega), \forall\omega$, i.e., a contour in the TF plane at the group delay.

As to phase-rate dominance, the most straightforward case to consider is a uniform filterbank constructed from Gaussian filters for which $Z_\tau = -\beta^2\tau$, $Z_\mu = \jmath\frac{\mathrm{d}\omega}{\mathrm{d}\mu}\tau$, $\boldsymbol{\phi}^T = \left[0\ \frac{\mathrm{d}\omega}{\mathrm{d}\mu}\tau\right]$ and $\mathbf{a}^T = \left[-\beta^2\tau\ 0\right]$. Using (22) and (24), the test for phase dominance becomes

$$p(\tau,\mu) = \frac{|\boldsymbol{\phi}^T(\tau,\mu)\mathbf{W}\,\mathbf{a}(\tau,\mu)|}{\|\mathbf{a}(\tau,\mu)\|^2} = \frac{1}{\beta^2}\frac{\mathrm{d}\omega}{\mathrm{d}\mu} > C_2. \tag{29}$$

Recalling that for uniform filter banks $\frac{\mathrm{d}\omega}{\mathrm{d}\mu} = \omega_{\max} - \omega_{\min}$, gives

$$\beta < \sqrt{\frac{\omega_{\max}-\omega_{\min}}{C_2}}.$$

In the following sub-section this result will be considered again and contrasted with a related result involving the response to a single phasor.

For a gammachirp prototype (27) and (28) become

$$Z_\tau = \beta\left\{\frac{n-1}{\beta\tau} - \frac{n+1}{n} + \jmath c\left\{\frac{1}{\beta\tau} - \frac{1}{n}\right\}\right\} \tag{30}$$

and

$$Z_\mu = \left\{\frac{\mathrm{d}\omega}{\mathrm{d}\mu}\frac{1}{\beta}\right\}\{n-\beta\tau\}\left\{\frac{\mathrm{d}\beta}{\mathrm{d}\omega} + \jmath\left\{\frac{\mathrm{d}\beta}{\mathrm{d}\omega}\frac{c}{n} - 1\right\}\right\} \tag{31}$$

respectively. Clearly both $\Re\{Z_\tau\}$ and $\Re\{Z_\mu\}$ and hence $\|\mathbf{a}(\tau,\mu)\|$ are not dependent on c. Fig. 1(a) illustrates the relationship between $p(\tau,\mu)$ and $\|\mathbf{a}(\tau,\mu)\|$ for a single filter in a filterbank where $\frac{\mathrm{d}\beta}{\mathrm{d}\omega} = \frac{1}{9}$, $\frac{\mathrm{d}\omega}{\mathrm{d}\mu}\frac{1}{\beta} = 35$ and $\beta = 712$ rad/s for 3 values of the filter parameter c. The stationary phase point is at $\beta\tau = n = 4$ whereas the peak magnitude of the impulse response occurs is at $\beta\tau = n-1 = 3$, cf. Fig. 1(b). As pointed out above, $\|\mathbf{a}(\tau,\mu)\|$ is identical for all values of c. On the other hand increasing the value of c increases the region of $\beta\tau$ where phase rate dominance can be achieved. Fig. 2 shows both the complete TF response and results for TFSPA for $n = 4$ and $c = 4$. The stationary phase points that satisfy (20) are indicated in red on Fig. 2(b). This stationary phase contour is not co-incident with the peak response or ridge but rather lies at the group delay of each filter as shown earlier. Fig. 2(b) is obtained by only plotting $|X(\tau,\omega)|$ at points in the TF plane where (26) is not satisfied. As expected from Fig. 1, the leading edge of the response has been removed. Reconstructions of the input waveform are shown in Fig. 3 using both full TF plane and TFSPA. In this case they are almost identical despite the removal of the leading edge of the response.

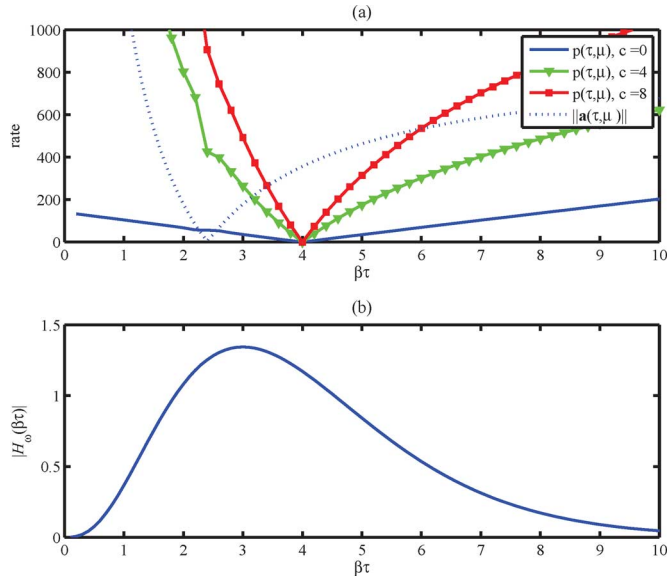

Fig. 1. Impulse response of gammachirp filter at 1 kHz: (a) comparison of $p(\tau,\mu)$ and $\|\mathbf{a}(\tau,\mu)\|$ at various values of chirp rate parameter c; (b) response.

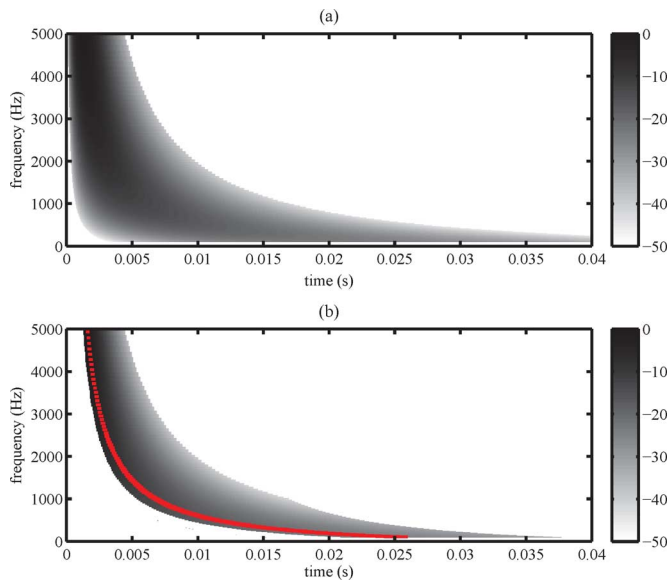

Fig. 2. Response a gammachirp cochlear-spaced filter bank to single impulse: (a) $|X(\tau,\omega)|$ in dB; (b) TFSPA—stationary phase points in red.

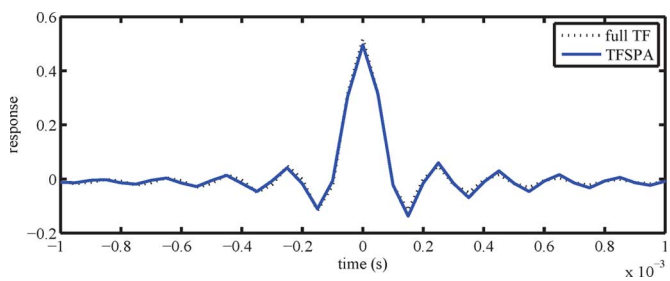

Fig. 3. Reconstruction of impulse using full TF plane of Fig. 2(a) and TFSPA of Fig. 2(b).

B. A Single Phasor

Apply a single tone of λ rad/s to the analysis filter bank i.e., $x(t) = u(t)e^{\jmath\lambda t}$, where $u(t)$ is the step function at the origin. In the steady state for $t \gg 0$ this gives

$$X(\tau,\omega) \approx \mathrm{e}^{\jmath\lambda\tau} H(\Omega)$$

where $\Omega = \frac{\lambda-\omega}{\beta}$, with $Z_\tau = \beta\{\eta + \jmath\Omega\}$ and

$$Z_\mu = \frac{\mathrm{d}\omega}{\mathrm{d}\mu}\frac{1}{\beta}\left\{\frac{d\beta}{d\omega}\{1+\eta\beta g(\omega)\} - \frac{\dot{H}(\Omega)}{H(\Omega)}\left\{1+\frac{d\beta}{d\omega}\Omega\right\} - \jmath\beta g(\omega)\right\} \quad (32)$$

Thus for all filters considered at any of the filter spacings considered the stationary phase point occurs when $\lambda = \omega$, $\forall\tau$ since, by definition, the group delay of the filter at ω is $g(\omega) = -\Im\{\dot{H}(0)/H(0)\}/\beta$. As to phase-rate dominance, the most straightforward case to consider is again a uniform filterbank constructed from Gaussian filters. For a Gaussian prototype, $g(\omega) = 0$, $\dot{H}(\Omega) = -\Omega H(\Omega)$ and $\eta = 0$. Thus $Z_\tau = \jmath\beta\Omega$ and

$$Z_\mu = \left\{\frac{\mathrm{d}\omega}{\mathrm{d}\mu}\frac{1}{\beta}\right\}\left\{\frac{d\beta}{d\omega} + \Omega\left\{1+\frac{\mathrm{d}\beta}{\mathrm{d}\omega}\Omega\right\}\right\}.$$

Noting that Z_τ is purely imaginary and Z_μ is purely real, $\|\mathbf{a}(\tau,\mu)\| = |Z_\mu|$ and $p(\tau,\mu) = \beta|\Omega|$, the inequality (22), reduces to

$$\beta|\Omega| > \left\{\frac{\mathrm{d}\omega}{\mathrm{d}\mu}\frac{1}{\beta}\right\}\left|\frac{d\beta}{d\omega} + \Omega\left\{1+\frac{\mathrm{d}\beta}{\mathrm{d}\omega}\Omega\right\}\right| \quad (33)$$

For uniformly spaced filters, this will be satisfied if $\beta > \{\frac{\mathrm{d}\omega}{\mathrm{d}\mu}\frac{1}{\beta}\}$. Note that this in direct contradiction to (29) and hence it is not possible to design a Gaussian filter bank that exhibits phase-rate dominance in response to both impulse and phasor inputs. For a *gammachirp* prototype, $g(\omega) = n/\beta$, $\dot{H}(\Omega)/H(\Omega) = \frac{-n\{\Omega+j\}-\jmath c\{\Omega+c/n\}}{1+\{\Omega+c/n\}^2}$ and $\eta = -1/n$. Thus

$$Z_\tau = \beta\left\{-\frac{1}{n} + \jmath\Omega\right\}$$

and

$$Z_\mu = \frac{\mathrm{d}\omega}{\mathrm{d}\mu}\frac{n\Omega}{\beta} \times \left\{\frac{\{\frac{\mathrm{d}\beta}{\mathrm{d}\omega}\Omega+1\} + \jmath\frac{\mathrm{d}\beta}{\mathrm{d}\omega} + \jmath\{\Omega+\frac{c}{n}\}\{\frac{c}{n}\frac{\mathrm{d}\beta}{\mathrm{d}\omega}-1\}}{1+\{\Omega+\frac{c}{n}\}^2}\right\}.$$

Fig. 4(a) illustrates the relationship between $p(\tau,\mu)$ and $\|\mathbf{a}(\tau,\mu)\|$ for a single filter in a filterbank described in Section V.A. The norm $\|\mathbf{a}(\tau,\mu)\|$ is approximately constant and virtually independent of c. The projection $p(\tau,\mu)$ is approximately linear in Ω and much less dependent on c than the response to an impulse. For these examples, (26) is satisfied and phase rate dominance is achieved outside the nominal bandwidth β of the filters, i.e., for $|\Omega| > \frac{1}{2}$. The implication is that the output of an analysis filter at ω rad/s, in response to a phasor at λ rad/s, contributes little to the output of the synthesis filter bank if $|\lambda-\omega| > \frac{\beta}{2}$. Further, (26) can be used to locate the vicinity of the peak in the response $|X(\tau,\omega)|$.

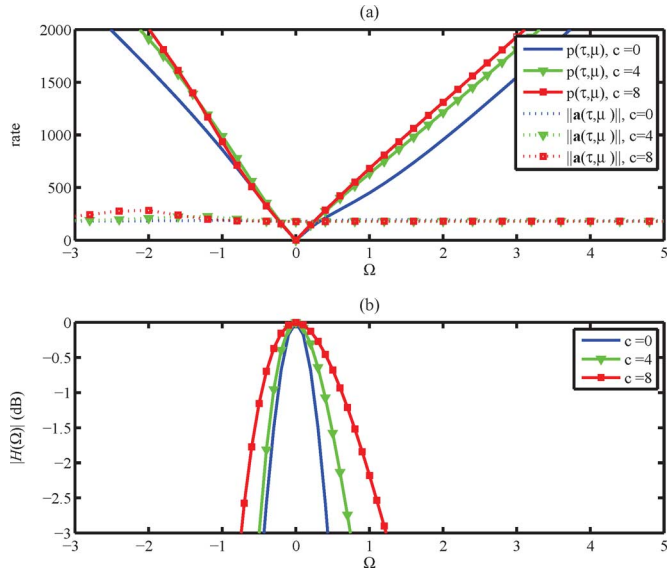

Fig. 4. Steady state response of gammachirp cochlear-spaced filter bank to 1 kHz tone: (a) rate; (b) gain.

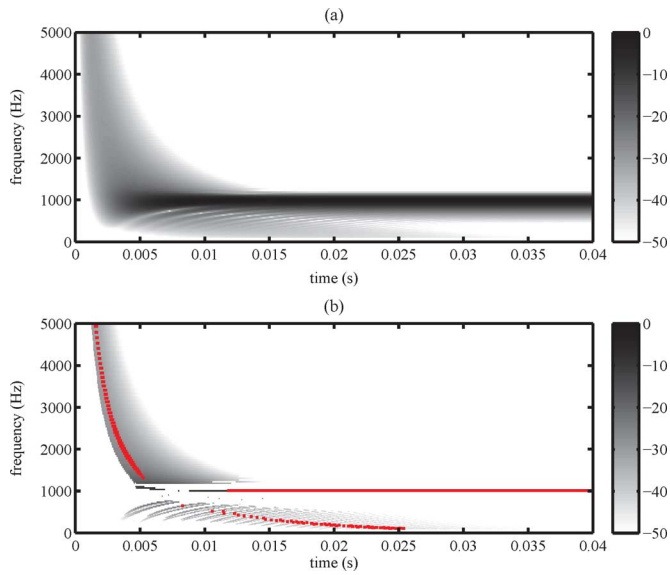

Fig. 5. TF response of a gammachirp cochlear-spaced filter bank to a 1 kHz tone applied at time zero: (a) $|X(\tau,\omega)|$ in dB; (b) TFSPA - stationary phase points in red.

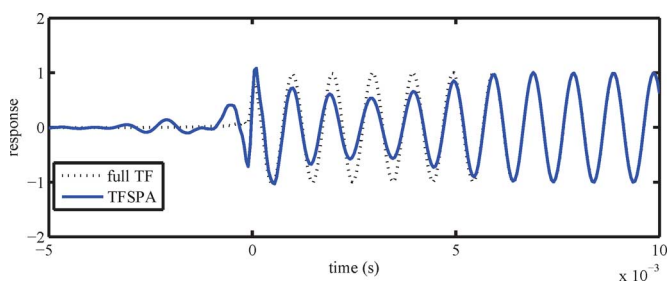

Fig. 6. Reconstruction of 1 kHz tone applied at time zero.

Fig. 5 shows both the complete TF response and the results for TFSPA. The stationary phase points are indicated in red on Fig. 5(b). The remaining values of $|X(\tau,\omega)|$ that are plotted are

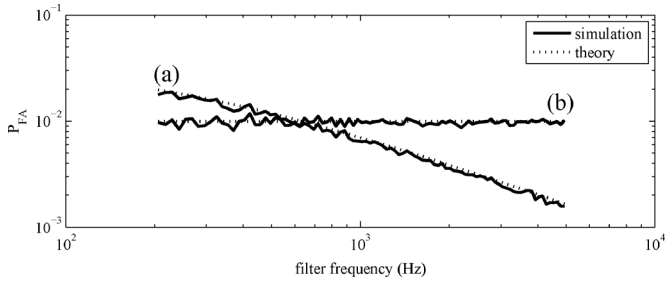

Fig. 7. Probability of false alarm for cochlear-spaced gammatone filter bank: theoretical and simulated performance of both (a) deterministic detector of (20) and (b) CFAR detector of (21).

at points in the TF plane where (26) is not satisfied. The steady state behavior is well predicted from the theoretical considerations above as illustrated in Fig. 4. The horizontal red line at 1000 Hz corresponds with the stationary phase point on Fig. 4. There is only a small region around $\Omega = 0$ in Fig. 4 where (26) is not satisfied. This region is just visible on Fig. 5 as a horizontal band in red at 1000 Hz. The transient behavior has similarities to that of the impulse response of Fig. 2(b). Specifically the stationary phase points, on Fig. 2(b) at the group delay, are present at most frequencies on the left hand side of Fig. 5(b) apart from a band of frequencies around the applied one. Fig. 6 shows the reconstruction achieved both with the whole TF plane and with TFSPA. The transient response of TFSPA introduces an amplitude modulation that dies away before matching the steady state response. This modulation is evident for $\tau < 7$ ms, the group delay of the filter, after which the stationary phase points at 1000 Hz is established.

In noisy environments, any detection system such as (20) or (21) will make type I errors [21]. In [19] a method for calculating the probability of false alarm of stationary phase point detectors such as (20) was presented. The related contribution proposed here is the CFAR detector of (21). Its performance is verified by simulation in Fig. 7. This detector is designed to have a false alarm rate of 10^{-2} at all frequencies of a cochlear-spaced gammatone filterbank. For reference the theoretical and measured performance of the deterministic detector of (20) is also included. The false alarm rate of the latter varies with frequency. These results confirm the validity of the independence assumption used in Appendix B to design the CFAR detector.

C. Linear Chirp

Consider a linear chirp with a rate of γ rad/s^2, i.e., $x(t) = u(t)\mathrm{e}^{\jmath\frac{\gamma}{2}t^2}$. The response to this chirp is $X(\tau,\omega) = \int_0^\tau \beta h(\beta t)\mathrm{e}^{\jmath\omega t}\mathrm{e}^{\jmath\frac{\gamma}{2}\{\tau-t\}^2}\mathrm{d}t$. In the steady state, when $\tau \gg \frac{1}{\beta}$ and $|h(\beta t)| \to 0$, this simplifies to

$$X(\tau,\omega) = \mathrm{e}^{\jmath\frac{\gamma}{2}\tau^2}\int_0^\infty \left\{\beta h(\beta t)\mathrm{e}^{-\jmath\gamma\tau t}\mathrm{e}^{\jmath\frac{\gamma}{2}t^2}\right\}\mathrm{e}^{\jmath\omega t}\mathrm{d}t.$$

This integral is intractable and is often approximated using the PSP. However it might appear imprudent to proceed with repeated application of that principle. The approach adopted here is to approximate the integral by more explicit means. First, assume that the instantaneous frequency, γt, of the quadratic phase term $\mathrm{e}^{\jmath\frac{\gamma}{2}t^2}$ is constant over the temporal extent of $h(\beta t)$. This approximation is valid if $|\gamma| \ll \beta^2$. Then, replace the quadratic phase term with a linear phase term whose frequency is identical to that of the quadratic phase term at a time equal to the group delay, i.e., $\mathrm{e}^{\jmath\frac{\gamma}{2}t^2} \approx \mathrm{e}^{\jmath\gamma g(\omega)t}$. This gives

$$X(\tau,\omega) = \mathrm{e}^{\jmath\frac{\gamma}{2}\tau^2}H\left(\frac{-\omega+\gamma\tau-\gamma g(\omega)}{\beta}\right).$$

This the envelope $|X(\tau,\omega)|$ has a peak of unity on the line $\tau = \frac{\omega}{\gamma} + g(\omega)$ in the TF plane. The stationary phase points, on the other hand, are defined as the solution to (18), which leads to

$$\gamma\tau + \Im\left\{\frac{\dot{H}\left(\frac{-\omega+\gamma\tau-\gamma g(\omega)}{\beta}\right)}{H\left(\frac{-\omega+\gamma\tau-\gamma g(\omega)}{\beta}\right)}\right\}\frac{\gamma}{\beta} = \omega \qquad (34)$$

and

$$\Im\left\{\frac{\dot{H}\left(\frac{-\omega+\gamma\tau-\gamma g(\omega)}{\beta}\right)}{H\left(\frac{-\omega+\gamma\tau-\gamma g(\omega)}{\beta}\right)}\right\}\left\{\frac{-1-\gamma\dot{g}(\omega)}{\beta}\right\} = g(\omega)$$

respectively, where $\dot{g}(\omega) = \frac{\mathrm{d}}{\mathrm{d}\omega}g(\omega)$. Combining these two expression to remove $\Im\{\dot{H}(\,\cdot\,)/H(\,\cdot\,)\}$ leads to the conclusion that the stationary phase points line on the line

$$\tau = \frac{\omega}{\gamma} + \frac{g(\omega)}{1+\gamma\dot{g}(\omega)} \qquad (35)$$

in the TF plane. For a Gaussian prototype, $g(\omega) = 0$, $\dot{g}(\omega) = 0$ and hence the stationary phase points are coincident with the peak in response. For gammatone and gammachirp prototypes, $g(\omega) = \frac{n}{\beta}$ and hence $\gamma\dot{g}(\omega) = -n\frac{\gamma}{\beta^2}\frac{\mathrm{d}\beta}{\mathrm{d}\omega}$. Then, since $\frac{\mathrm{d}\beta}{\mathrm{d}\omega} \ll 1$, $|\gamma| \ll \beta^2$ and typically $n < 10$, the denominator $1+\gamma\dot{g}(\omega)$ is approximately one. Thus the stationary phase points are approximately co-incident with the peak in the response. When the chirp rate is low, i.e., provided $|\gamma| \ll \beta^2$, the behavior of the $p(\tau,\mu)$ and $\|\mathbf{a}(\tau,\mu)\|$ can be inferred from Fig. 4. There will be a band in the TF plane around the stationary phase line where the inequality (26) is not satisfied.

Fig. 8 illustrates these points. The applied signal is the sum of two linear chirps. The first is a down-chirp ($\gamma = -10\times10^4$ rad/s/s) which can be observed in Fig. 8(a) as a continuous ridge from top left to bottom right. A second lower amplitude up-chirp ($\gamma = 8\times10^4$ rad/s/s) produces a ridge from bottom left to upper right. The down chirp produces a series of stationary phase points (in red) from top left to bottom right. The position of these points is well predicted by (35), shown as red dots, apart from bottom right where the inequality, $|\gamma| \ll \beta^2$, does not hold. As expected there is a region of the TF plane around the stationary phase lines where (26) in not satisfied. The lower amplitude up-chirp is depicted in a similar manner (including stationary phase points) apart from the region where the two chirps cross. In this region the larger amplitude chirps hides the trajectory of the lower amplitude chirp. This is a form of simultaneous masking that will be explored more thoroughly in Section VI. Finally there is a low amplitude artifact that lies between the two chirps. This artifact does not contain stationary phase points and could be ignored on that basis in applications where signal analysis was the primary objective. The outputs from the synthesis filter banks are shown in Fig. 9 at the point where the two frequencies cross. TFSPA produces some distortion in the reconstruction when compared with that produced using the whole TF plane.

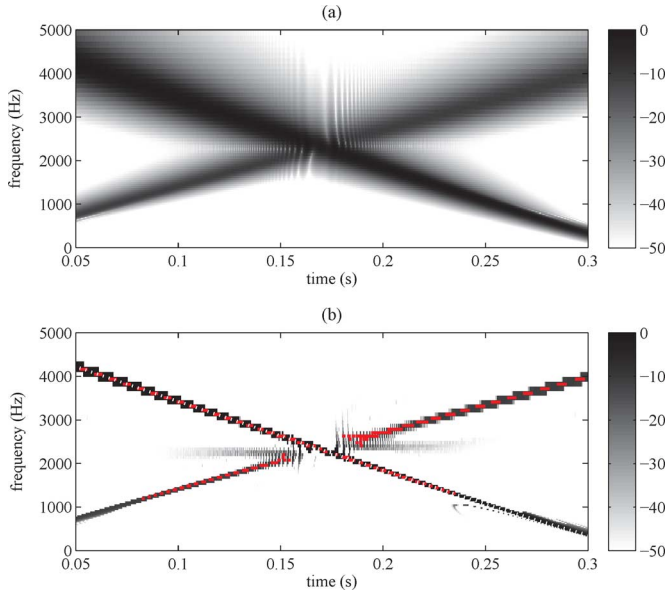

Fig. 8. TF response of a gammatone cochlear-spaced filter bank to two linear chirps (a) $\lfloor X(\tau,\omega)\rfloor$ in dB; (b) TFSPA—stationary phase points in red—equation (35) for downchirp (white dots against red and black).

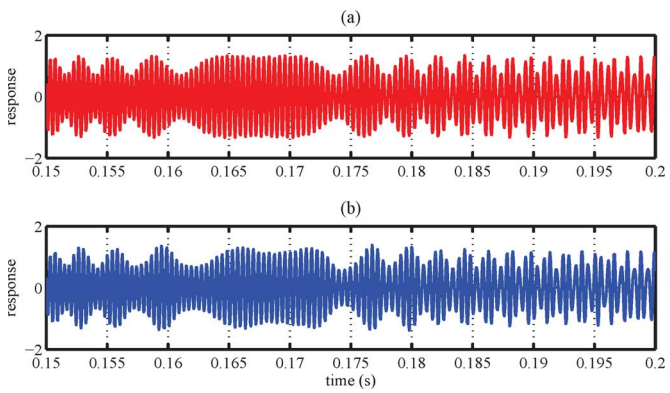

Fig. 9. Reconstruction of chirps in the cross-over region using: (a) full TF plane of Fig. 8(a); (b) TFSPA of Fig. 8(b).

D. A Decaying Phasor and Voiced Speech

Voiced speech can be viewed as a signal of the form $\sum_{i=1}^{N}\sum_{n} A_i u(t-nT)\mathrm{e}^{\lambda_i\{t-nT\}}$ where T is the pitch period and $\{\lambda_i\}_{i=1}^{N}$ are the N complex formant frequencies with N centre frequencies $\{\Im\{\lambda_i\}\}_{i=1}^{N}$ and time constants $\{-\Re\{\lambda_i\}\}_{i=1}^{N}$; $\{A_i\}_{i=1}^{N}$ are the complex amplitudes of this atomic decomposition. Because of the relationship to speech it is informative to consider the stationary phase points associated with a single atom of this decomposition, i.e., when $x(t)=u(t)\mathrm{e}^{\lambda t}$ with a single complex frequency λ, $\Re\{\lambda\}<0$. This gives

$$X(\tau,\omega)=\beta\mathrm{e}^{\lambda\tau}\int_{-\infty}^{\tau}h(\beta t)\mathrm{e}^{\jmath\omega t}\mathrm{e}^{-\lambda t}\mathrm{d}t.$$

To proceed further consider the case where $h(t)\in\mathbb{R}$ and thus restrict consideration to Gaussian and gammatone prototypes. The normalized time derivative is

$$\frac{\frac{\partial}{\partial\tau}X(\tau,\omega)}{X(\tau,\omega)}=\frac{h(\beta\tau)\mathrm{e}^{\jmath\omega\tau}\mathrm{e}^{-\lambda\tau}}{\int_{-\infty}^{\tau}h(\beta t)\mathrm{e}^{\jmath\omega t}\mathrm{e}^{-\lambda t}\mathrm{d}t}+\lambda.$$

Note that since $h(t)$ is real, the first term on the right hand side is also real if $\Im(\lambda)=\omega$ at which point: $\Im\{Z_\tau\}=0$, cf. (12). In a similar manner

$$\frac{\frac{\partial}{\partial\omega}X(\tau,\omega)}{X(\tau,\omega)}=\frac{\jmath\int_{-\infty}^{\tau}h(\beta t)\mathrm{e}^{\jmath\omega t}\mathrm{e}^{-\lambda t}t\mathrm{d}t}{\int_{-\infty}^{\tau}h(\beta t)\mathrm{e}^{\jmath\omega t}\mathrm{e}^{-\lambda t}\mathrm{d}t}$$

which is purely imaginary for Gaussian and gammatone prototypes. Thus setting $\Im\{Z_\mu\}=0$ for $\Im\{\lambda\}=\omega$ gives

$$\int_{-\infty}^{\tau}h(\beta t)\mathrm{e}^{-\Re(\lambda)t}\{t-g(\omega)\}\mathrm{d}t=0, \tag{36}$$

cf. (13). While it is unlikely that closed form solutions of (36) for τ can be found, nevertheless it is still possible to infer something about the nature of the solutions. Because $h(\beta t)\mathrm{e}^{-\Re(\lambda)t}>0,\forall t$ and $\{t-g(\omega)\}<0,\forall t<g(\omega)$, no solution exists for $\tau<g(\omega)$. For $\tau\geq g(\omega)$ the integral can be split into two terms, i.e., $\int_{-\infty}^{\tau}\mathrm{d}t=\int_{-\infty}^{g(\omega)}\mathrm{d}t+\int_{g(\omega)}^{\tau}\mathrm{d}t$. The first term is a negative constant. The second is a nonnegative monotonically increasing function of τ because $h(\beta t)\mathrm{e}^{-\Re(\lambda)t}\{t-g(\omega)\}\geq 0,\forall\tau\geq g(\omega)$. Thus, if a solution exists it will be the only solution. Such a solution will obviously be dependent on the value of the bandwidth β of the analysis filter for which $\omega=\Im(\lambda)$ and the time constant $-\Re(\lambda)$ of the applied pulse. Thus it may be possible to infer both the frequency and time constant of the applied pulse from the position of the stationary phase point.

To further explore the response of TFSPA to decaying phasors and to provide an example of its application, a short segment of voiced speech is used. In this case the signal $x(t)$ is the first utterance of the phoneme /ax/ from the first record of the Timit Database at [23]. The voice is a US male speaker sampled at 16 kHz. A gammatone cochlear-spaced filter bank is used for the analysis. For this experiment the spacing of the filters in frequency is such that there are 16 filters per ERB. This is commensurate with the frequency accuracy of the human auditory system which is known to be a fraction of the ERB [18]. To allow for the possibility of noise in the recorded data the detector of (21) is used with a false alarm rate of 1/500. From Fig. 10(a) it is clear that the signal is quasi-periodic in time and consists of a number of formant frequencies [18]: a formant just above 3 kHz which changes in frequency with time; a formant just below 2 kHz; a formant just below 1.5 kHz; up to 4 resolvable frequency components below 500 Hz. The corresponding result for TFSPA is shown in Fig. 10(b). The representation is notably sparser. The formant structure is clearer, particularly over the frequency range 1–3.5 kHz. As predicted from (36), there is a single stationary phase point in response to the onset of each decaying phasor in the speech waveform. This is particularly evident for the formant just above 3 kHz but is also evident for the formants just below 2 kHz and below 1.5 kHz. Another feature is predicted by the theoretical consideration in Section V-A and is also visible in Figs. 2(b) and 5(b). Specifically there is a line of stationary phase points one group delay after the start of each period of the applied waveform. This is visible in the band 500–1500 Hz, particularly on the left of Fig. 10(b).

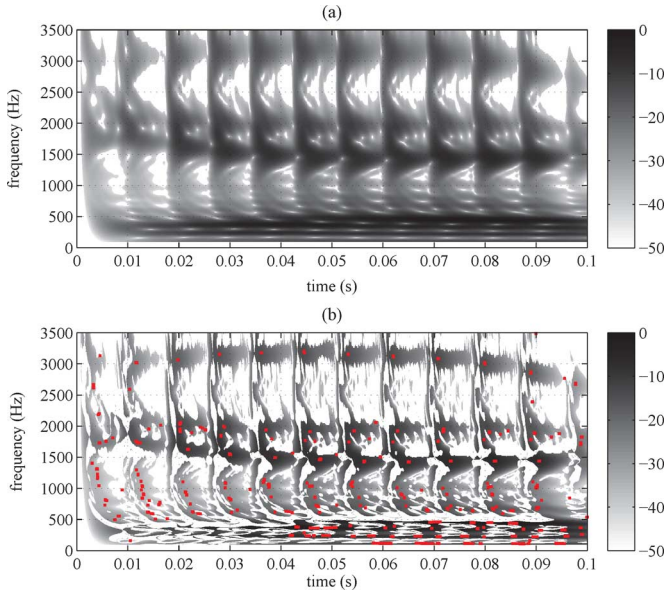

Fig. 10. TF response of a gammatone cochlear-spaced filter bank to phoneme /ax/ from record 1 of the Timit Database [23]: (a) $|X(\tau,\omega)|$ in dB; (b) TFSPA - stationary phase points in red.

VI. Simultaneous Masking

Simultaneous audio masking is the well studied process whereby the presence of one tone prevents the detection of a second tone [18]. In this section it is shown that the use of the test (20) leads to a form of simultaneous masking. In particular, if the test (20) is used to indicate the presence or otherwise of a tonal component at ω, the amount by which that test is de-sensitized by the presence of a second tone at ω_1 is dependent on: (i) the frequency separation of the tones $\{\omega_1-\omega\}$; (ii) the relative amplitude $|A_1|$ of the tones; (iii) the magnitude frequency response $|H(\frac{\{\omega_1-\omega\}}{\beta})|$ of the filter at ω. Consider a signal $x(t)=\mathrm{e}^{\jmath\omega t}+A_1\mathrm{e}^{\jmath\omega_1 t}$ made up of two tones at ω and ω_1, where A_1, the relative amplitude, is positive real. The steady-state response of the analysis filter at ω is

$$X(\tau,\omega)=\mathrm{e}^{\jmath\omega\tau}+A_1H_1\mathrm{e}^{\jmath\omega_1\tau}$$

where $H_1=H(\Omega_1)$ and $\Omega_1=\{\omega_1-\omega\}/\beta$ is the normalized frequency separation with derivative $\frac{\mathrm{d}\Omega_1}{\mathrm{d}\omega}=-\frac{1}{\beta}\{\Omega_1\frac{\mathrm{d}\beta}{\mathrm{d}\omega}+1\}$. The time derivative follows:

$$\frac{\partial}{\partial\tau}X(\tau,\omega)=\mathrm{e}^{\jmath\omega\tau}\jmath\omega+A_1H_1\mathrm{e}^{\jmath\omega_1\tau}\jmath\omega_1$$

as does the frequency derivative:

$$\frac{\partial}{\partial\omega}X(\tau,\omega)=-\frac{1}{\beta}\dot{H}_0\mathrm{e}^{\jmath\omega\tau}+A_1\frac{\mathrm{d}\Omega_1}{\mathrm{d}\omega}\dot{H}_1\mathrm{e}^{\jmath\omega_1\tau}$$

where $H_0=H(0)$. It is convenient to rewrite (14) as

$$\Im\{Z_\tau\}=\frac{\Im\left\{X^*(\tau,\omega)\frac{\partial}{\partial\tau}X(\tau,\omega)\right\}-\omega|X(\tau,\omega)|^2}{|X(\tau,\omega)|^2} \quad (37)$$

and (15) as

$$\Im\{Z_\mu\}=\left\{\frac{\Im\left\{X^*(\tau,\omega)\frac{\partial}{\partial\omega}X(\tau,\omega)\right\}-g(\omega)|X(\tau,\omega)|^2}{|X(\tau,\omega)|^2}\right\}\frac{d\omega}{d\mu} \quad (38)$$

with denominator

$$|X(\tau,\omega)|^2=1+|A_1H_1|^2+2|A_1H_1|\cos(\theta) \quad (39)$$

and numerator of (37) as

$$\Im\left\{X^*(\tau,\omega)\frac{\partial}{\partial\tau}X(\tau,\omega)\right\}-\omega|X(\tau,\omega)|^2 = |A_1H_1|\beta\Omega_1\left\{|A_1H_1|+\cos(\theta)\right\}. \quad (40)$$

where $\theta=\{\omega-\omega_1\}\tau-\phi$ and $\phi=\angle\{A_1H_1\}$. The angle variable θ indicates the position in time with respect to one period of the beat frequency $\{\omega-\omega_1\}$. In (38), the term $\Im\{X^*(\tau,\omega)\frac{\partial}{\partial\omega}X(\tau,\omega)\}$ is

$$\begin{aligned}\Im\{X^*(\tau,\omega)&\frac{\partial}{\partial\omega}X(\tau,\omega)\}\\ &=\Im\left\{-\frac{1}{\beta}\dot{H}_0+\frac{\mathrm{d}\Omega_1}{\mathrm{d}\omega}|A_1|^2H_1^*\dot{H}_1-\frac{1}{\beta}A_1^*H_1^*\dot{H}_0\mathrm{e}^{\jmath\{\omega-\omega_1\}\tau}\right\}\\ &\quad+\Im\left\{\frac{\mathrm{d}\Omega_1}{\mathrm{d}\omega}A_1\dot{H}_1\mathrm{e}^{\jmath\{\omega_1-\omega\}\tau}\right\}. \quad (41)\end{aligned}$$

Consider a *gammatone* prototype for which $g(\omega)=-n/\beta$, $H(\Omega)=\frac{1}{\{1+\jmath\Omega\}^n}$ and $\frac{\dot{H}(\Omega)}{H(\Omega)}=\frac{-n\{\Omega+j\}}{1+\Omega^2}$. Hence $\Im\{\dot{H}_0\}=-n$, $\Im\{H_1^*\dot{H}_1\}=\frac{-n|H_1|^2}{1+\Omega_1^2}$ and $H_1^*\dot{H}_0=-\jmath nH_1^*$ from which the normalized frequency derivative is obtained using (41).

$$\begin{aligned}\Im\{X^*(\tau,\omega)&\frac{\partial}{\partial\omega}X(\tau,\omega)\}-g(\omega)|X(\tau,\omega)|^2\\ &=-\frac{n}{\beta}\frac{\Omega_1}{1+\Omega_1^2}|A_1H_1|\left\{\Omega_1-\frac{\mathrm{d}\beta}{\mathrm{d}\omega}\right\}\left\{|A_1H_1|+\cos(\theta)\right\}\\ &\quad-\frac{n}{\beta}\frac{\Omega_1}{1+\Omega_1^2}|A_1H_1|\left\{\Omega_1\frac{\mathrm{d}\beta}{\mathrm{d}\omega}+1\right\}\sin(\theta) \quad (42)\end{aligned}$$

After some manipulation this yields (47), where

$$\sin(\psi)=\frac{\Omega_1-\frac{\mathrm{d}\beta}{\mathrm{d}\omega}}{\sqrt{1+\left\{\frac{\mathrm{d}\beta}{\mathrm{d}\omega}\right\}^2}\sqrt{1+\Omega_1^2}} \quad (43)$$

$$\cos(\psi)=\frac{\Omega_1\frac{\mathrm{d}\beta}{\mathrm{d}\omega}+1}{\sqrt{1+\left\{\frac{\mathrm{d}\beta}{\mathrm{d}\omega}\right\}^2}\sqrt{1+\Omega_1^2}}. \quad (44)$$

The numerator of (47) is the sum of two oscillatory terms in θ. Assuming that the filterbank is such that $\beta^2\gg\{\frac{d\omega}{d\mu}\frac{n}{\beta}\}^2\{1+\{\frac{\mathrm{d}\beta}{\mathrm{d}\omega}\}^2\}$, then the first oscillatory term, from $\Im\{Z_\tau\}$, is larger than the second term, from $\Im\{Z_\mu\}$. To obtain an approximation to the norm, first consider $|A_1H_1|\ll 1$, in which case the denominator of (47) is a constant. For $|\Omega_1|\gg 1$, $|\sin(\psi)|\approx 1$, $|\cos(\psi)|\approx\frac{\mathrm{d}\beta}{\mathrm{d}\omega}$ (i.e., small and finite) and $\frac{\mathrm{d}\beta}{\mathrm{d}\omega}\approx 0$. Thus the numerator is the sum of two oscillatory terms proportional to $\{|A_1H_1|+\cos(\theta)\}^2$. Thus the minima of the norm are, approximately, at $\cos(\theta)=-A_1H_1$. For $\Omega_1\ll 1$, $|\cos(\psi)|\approx 1$

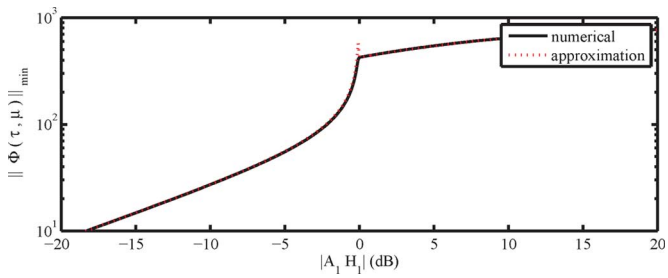

Fig. 11. Numerical and approximate evaluation of minimum of phase rate norm of cochlear spaced gammatone filter: filter frequency 1 kHz and $\Omega_1 = 1$.

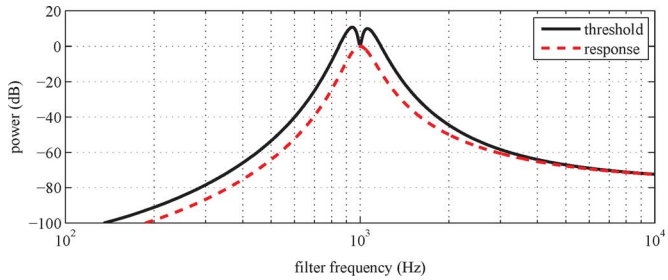

Fig. 12. Detection threshold of gammatone cochlear-spaced filter bank for $C_1 = 10$; 0 dB masking tone at 1 kHz; response of filterbank to masker also shown for reference.

and $|\sin(\psi)| \approx 0$ and hence the two oscillatory terms are approximately out of phase and thus the minima of the numerator can again be found at the zeros of the larger term, i.e., $\cos(\theta) = -A_1 H_1$. Substitution in (47) yields

$$\|\mathbf{\Phi}(\tau,\mu)\|_{\min} \approx \frac{\mathrm{d}\omega}{\mathrm{d}\mu}\frac{n}{\beta}\frac{|\Omega_1|}{1+\Omega_1^2}\frac{|A_1H_1|}{\sqrt{1-|A_1H_1|^2}} \tag{45}$$

For $|A_1H_1| \gg 1$ the first term does not go to zero. Given that this time derivative term much larger than the frequency derivative term, an approximation to the norm is obtained by assuming that the latter term is negligible and hence

$$\|\mathbf{\Phi}(\tau,\mu)\| \approx \frac{\beta\Omega_1|A_1H_1|\{|A_1H_1|+\cos(\theta)\}}{1+|A_1H_1|^2+2|A_1H_1|\cos(\theta)}.$$

This function of angle has minima at $\cos(\theta) = 1$, which are:

$$\|\mathbf{\Phi}(\tau,\mu)\|_{\min} \approx \frac{\beta|\Omega_1|A_1H_1|}{|A_1H_1|+1} \tag{46}$$

Together (45) and (46) approximate the behavior of the minima of the norm $\|\mathbf{\Phi}(\tau,\mu)\|$ as a function of the frequency separation Ω_1 and the amplitude of the second tone as observed at the output to the filter $|A_1H_1|$. When two tones are present the norm will not go to zero. However, despite that, the test (20) may be satisfied for a given threshold C_1 twice per period of the beat frequency at $\cos(\theta) = -|A_1H_1|$ and thus indicate that a region of stationary phase is present. It is also evident from (45) and (46) that the behavior of the norm changes significantly at the point where $|A_1H_1| = 1$. See (47) at the bottom of the page.

Fig. 11 illustrates $\|\mathbf{\Phi}(\tau,\mu)\|_{\min}$ as a function of $|A_1H_1|$ for a particular filter and a frequency separation, $\Omega_1 = 1$. The minimum value of the norm is evaluated both by numerical minimization of (47) and using the approximations of (45) and (46). There is a clear discontinuity in the first derivative at $|A_1H_1| = 1$ as might be expected from the development of the approximations. The approximations are generally a good fit to the minima evaluated numerically. For this example the most significant errors can be observed around $|A_1H_1| = 1$. It is also clear from the numerical evaluation that the minimum of the norm is a montotonically increasing function of $|A_1H_1|$ and as such is invertible. Thus if a particular value of the threshold C_1 in (20) is chosen to detect stationary phase points a corresponding value of A_1H_1 can be calculated that will just achieve a minimum of C_1. Together (45) and (46) are not guaranteed to provide a montonically increasing function. However minor adjustments in the vicinity of the discontinuity can overcome this and together they provide a simple means of inverting the function. Fig. 12 illustrates how adoption of the stationary phase test (20) leads to a simultaneous masking effect. A 0 dB masking tone at 1 kHz is introduced that corresponds to the interferer at ω_1 rad/s. For each filter frequency ω rad/s, (45) and (46) are used to calculate the amplitude of a tonal component at ω rad/s that would produce a value of minimum norm equal to the threshold value of C_1. This amplitude is converted to decibels to give the detection threshold. For reference, the response of each of the filters, i.e., $|X(\tau,\omega)|$, to the masking tone alone is shown. When $\omega_1 = \omega$ a single tone is present and the results of Sub-Section V.B apply. The norm will go to zero at ω_1 and thus the test will be satisfied. There is a region around ω_1 where the detection threshold increases as $|\omega - \omega_1|$ increases. However outwith that region the detection threshold follows the same trends as $|X(\tau,\omega)|$ and, as such, this masking effect is not symmetrical and affects filters above ω_1 more than below it.

VII. Conclusion

The starting point for this paper was an examination of the application of the PSP to non-asymptotic integrals in general and TF synthesis in particular. The conclusion was that only one aspect of the PSP, location of stationary phase points, is required. The second aspect, approximation of the integral through use of the second derivative of the phase of the integrand, is only needed when closed form expressions that approximate the integral are required. When this requirement is removed, the second aspect can be replaced with a test for phase rate dominance. Regions of the TF plane that pass the test and do not contain stationary phase points contribute little or nothing to the final output. Analysis values that lie in these regions can thus be set

$$\|\mathbf{\Phi}(\tau,\mu)\|^2 \propto \frac{\beta^2\{|A_1H_1|+\cos(\theta)\}^2 + \left\{\frac{d\omega}{d\mu}\frac{n}{\beta}\right\}^2\frac{1+\left\{\frac{\mathrm{d}\beta}{\mathrm{d}\omega}\right\}^2}{1+\Omega_1^2}\{\sin(\psi)\{|A_1H_1|+\cos(\theta)\}+\cos(\psi)\sin(\theta)\}^2}{\{1+|A_1H_1|^2+2|A_1H_1|\cos(\theta)\}^2} \tag{47}$$

to zero. In regions of the TF plane that fail the test or in the vicinity of stationary phase points, synthesis is performed in the usual way. In re-examining the application of the PSP to the TF synthesis integral, a new interpretation of the location parameters associated with the synthesis filters leads to: (i) tests for locating stationary phase points in the TF plane; (ii) a test for phase rate dominance in that plane. With this formulation the stationary phase regions of several elementary signals have been identified theoretically and it has been shown that sparse reconstructions of tones and chirps are possible. An analysis of the TF phase rate characteristics for the case of two simultaneous tones predicts and quantifies a form of simultaneous masking similar to that which characterizes the auditory system.

Appendix A

Consider a complex function $f(\mathbf{v}) = a(\mathbf{v})\mathrm{e}^{\jmath b(\mathbf{v})}$ of a vector $\mathbf{v} = [v_1\ v_2]^{\mathrm{T}}$ with $a, b, v_1, v_2 \in \mathbb{R}$. The gradient vector is $\boldsymbol{\nabla}_f \triangleq [\frac{\partial f}{\partial v_1}\ \frac{\partial f}{\partial v_2}]^{\mathrm{T}}$. Thus

$$\frac{\boldsymbol{\nabla}_f(\mathbf{v})}{f(\mathbf{v})} \triangleq \boldsymbol{\nabla}_a(\mathbf{v}) + \jmath\boldsymbol{\nabla}_b(\mathbf{v})$$

where $\boldsymbol{\nabla}_a(\mathbf{v}) = \frac{1}{a(\mathbf{v})}\left[\frac{\partial a(\mathbf{v})}{\partial v_1}\ \frac{\partial a(\mathbf{v})}{\partial v_2}\right]^T$ and $\boldsymbol{\nabla}_b(\mathbf{v}) = \left[\frac{\partial b(\mathbf{v})}{\partial v_1}\ \frac{\partial b(\mathbf{v})}{\partial v_2}\right]^T$. The integral of the function around a point $\mathbf{v}_0$ over an interval $S = \{\mathbf{v} : \|\mathbf{v}\| < r\}$ with radius r is $I_S(\mathbf{v}_0) = \int_S f(\mathbf{v}_0 + \mathbf{v})\mathrm{d}\mathbf{v}$. Assuming that the function is well approximated by its first order Taylor series over this interval and making a change of variable $\mathbf{v} = \mathbf{Q}\mathbf{u}$, where $\mathbf{Q}$ is a orthonormal rotation matrix chosen such that $\mathbf{Q}^T\boldsymbol{\nabla}_a(\mathbf{v}_0) = [\|\boldsymbol{\nabla}_a(\mathbf{v}_0)\|\ 0]^T$ and $\mathbf{Q}^T\boldsymbol{\nabla}_b(\mathbf{v}_0) \triangleq [\nabla_{b1}(\mathbf{v}_0)\ \nabla_{b2}(\mathbf{v}_0)]^T$, gives

$$\begin{aligned} I_S(\mathbf{v}_0) &\approx f(\mathbf{v}_0)\int_S \left\{1 + \mathbf{u}^T\mathbf{Q}^T\boldsymbol{\nabla}_a(\mathbf{v}_0) + \jmath\mathbf{u}^T\mathbf{Q}^T\boldsymbol{\nabla}_b(\mathbf{v}_0)\right\}\mathrm{d}\mathbf{u} \\ &= f(\mathbf{v}_0)\left\{\iint_S \mathrm{d}u_1\mathrm{d}u_2 \right. \\ &\qquad + \{\|\boldsymbol{\nabla}_a(\mathbf{v}_0)\| + \jmath\nabla_{b1}(\mathbf{v}_0)\}\iint_S u_1\mathrm{d}u_1\mathrm{d}u_2 \\ &\qquad \left. +\jmath\nabla_{b2}(\mathbf{v}_0)\iint_S u_2\mathrm{d}u_1\mathrm{d}u_2\right\}. \end{aligned}$$

Given that the interval is circular $\iint_S u_1\mathrm{d}u_1\mathrm{d}u_2 = \iint_S u_2\mathrm{d}u_2\mathrm{d}u_2$, the integral becomes

$$\begin{aligned} I_S(\mathbf{v}_0) \approx f(\mathbf{v}_0)\left\{\iint_S \mathrm{d}u_1\mathrm{d}u_2 + \{\|\boldsymbol{\nabla}_a(\mathbf{v}_0)\| \right. \\ \left. +\jmath\{\nabla_{b1}(\mathbf{v}_0) + \nabla_{b2}(\mathbf{v}_0)\}\}\iint_S u_1\mathrm{d}u_1\mathrm{d}u_2\right\} \end{aligned}$$

By definition, the rotation matrix is given by, $\mathbf{Q}^T = \begin{bmatrix}\cos(\theta) & \sin(\theta)\\ -\sin(\theta) & \cos(\theta)\end{bmatrix}$, where $\begin{bmatrix}\cos(\theta)\\ \sin(\theta)\end{bmatrix} = \frac{\boldsymbol{\nabla}_a(\mathbf{v}_0)}{\|\boldsymbol{\nabla}_a(\mathbf{v}_0)\|}$. Thus the integral over the interval will be dominated by the phase variations if $|\nabla_{b1}(\mathbf{v}_0) + \nabla_{b2}(\mathbf{v}_0)| \gg \|\boldsymbol{\nabla}_a(\mathbf{v}_0)\|$, where

$$\begin{aligned} \nabla_{b1}(\mathbf{v}_0) + \nabla_{b2}(\mathbf{v}_0) &= [1\quad 1]\mathbf{Q}^T\boldsymbol{\nabla}_b(\mathbf{v}_0) \\ &= \frac{\boldsymbol{\nabla}_a^T(\mathbf{v}_0)}{\|\boldsymbol{\nabla}_a(\mathbf{v}_0)\|}\mathbf{W}\boldsymbol{\nabla}_b(\mathbf{v}_0) \end{aligned}$$

and $\mathbf{W} = \begin{bmatrix}1 & 1\\ -1 & 1\end{bmatrix}$. Therefore a test for first order phase rate dominance is

$$\frac{|\boldsymbol{\nabla}_a^T(\mathbf{v}_0)\mathbf{W}\boldsymbol{\nabla}_b(\mathbf{v}_0)|}{\|\boldsymbol{\nabla}_a(\mathbf{v}_0)\|} \gg \|\boldsymbol{\nabla}_a(\mathbf{v}_0)\|. \tag{48}$$

Appendix B

Under the null hypothesis, no signal of interest is present and $x(t)$ is zero-mean complex white Gaussian noise with variance σ_n^2. Consequently $X(\tau,\omega)$, $\frac{\partial}{\partial\omega}X(\tau,\omega)$ and $\frac{\partial}{\partial\tau}X(\tau,\omega)$ are zero-mean complex Gaussian random variables whose variances and cross-correlations can be evaluated using the filter responses $h_\omega(\tau)$, $\frac{\partial}{\partial\omega}h_\omega(\tau)$ and $\frac{\partial}{\partial\tau}h_\omega(\tau)$ and the input noise variance σ_n^2. In [19] (using [24]), closed form expressions for the PDF, $f(z)$, and the CDF, $F(z)$, of the ratio $z = \Im\{\frac{a}{b}\}$ are given where a and b are jointly-Gaussian zero-mean complex random variables with variances σ_a^2 and σ_b^2 respectively and with cross correlation coefficient $\rho = \frac{E[ab^*]}{\sigma_a\sigma_b}$. These results are similar to those obtained in [25] where a closed form expression for the PDF of the reassignment vector is obtained for Gaussian noise and a real symmetric window $h(t)$. Other aspects of the effects of noise on reassignment are addressed in [26] and [27]. From [24], the CDF has the form

$$F(z; z_0, s) = \frac{1}{2}\left\{1 + \left\{\frac{s}{z - z_0}\right\}^2\right\}^{-1/2} + \frac{1}{2}$$

where $z_0 = -\Im\{\rho\}\frac{\sigma_a}{\sigma_b}$ is the location parameter and $s = \frac{\sigma_a}{\sigma_b}\sqrt{1 - |\rho|^2}$ is the spread parameter. Expressions for the PDF and CDF of both $\Im\{Z_\tau\}$ and $\Im\{Z_\mu\}$ can be constructed from these using the appropriate parameters, i.e., $z_{0\tau}$ and s_τ for $\Im\{Z_\tau\}$ and $z_{0\mu}$ and s_μ for $\Im\{Z_\mu\}$. As in [25], the parameters of this distribution are functions of the ratio $\frac{\sigma_a}{\sigma_b}$. Therefore the distribution is not dependent on the noise variance σ_n^2 and knowledge of that variance is not required in specifying the false alarm rate of a detector. Assuming that $\Im\{Z_\tau\}$ and $\Im\{Z_\mu\}$ are independent[1], the probability of false alarm P_{fa} for the test (21) can be expressed in terms of the two positive thresholds $T_\tau(\mu)$ and $T_\mu(\mu)$, i.e., $P_{fa}(T_\tau, T_\mu) = P_\tau P_\mu$ where

$$\begin{aligned} P_\tau &= F(T_\tau; z_{0\tau}, s_\tau) - F(-T_\tau; z_{0\tau}, s_\tau) \\ P_\mu &= F(T_\mu; z_{0\mu}, s_\mu) - F(-T_\mu; z_{0\mu}, s_\mu) \end{aligned}$$

To eliminate one of the variables and equalize the sensitivity of the test in the two dimensions of time and frequency, set $P_\tau = P_\mu$. Thus $P_{fa}(T_\tau, T_\mu)$ can be inverted to specify all thresholds for a particular filterbank design.

[1]This assumption is justified empirically in [19] for gammatone filters by using it to reliably predict the false alarm rate of the simple detector of (20).

ACKNOWLEDGMENT

The author would like to thank the anonymous reviewers for their comments. Their suggestions helped improve this article significantly.

REFERENCES

[1] R. Wong, *Asymptotic Approximations of Integrals*, ser. Classics in Applied Mathematics. Philadelphia, PA, USA: SIAM, 2001, vol. 34.
[2] M. Soumekh, *Synthetic Aperture Radar Signal Processing*. New York, NY, USA: Wiley, 1999.
[3] E. Chassande-Mottin and P. Flandrin, "On the stationary phase approximation of chirp spectra," in *Proc. IEEE-SP Int. Symp. Time-Frequency Time-Scale Anal.*, 6–9, 1998, pp. 117–120.
[4] S. Mallat, *A Wavelet Tour of Signal Processing: The Sparse Way*, 3rd ed. New York, NY, USA: Academic, 2009.
[5] S. Mandal, S. Zhak, and R. Sarpeshkar, "A bio-inspired active radio-frequency silicon cochlea," *IEEE J. Solid-State Circuits*, vol. 44, no. 6, pp. 1814–1828, Jun. 2009.
[6] M. L. Jepsen, S. D. Ewert, and T. Dau, "A computational model of human auditory signal processing and perception," *J. Acoust. Soc. Amer.*, vol. 124, no. 1, pp. 422–438, 2008.
[7] N. Delprat, B. Escudie, P. Guillemain, R. Kronland-Martinet, P. Tchamitchian, and B. Torresani, "Asymptotic wavelet and Gabor analysis: Extraction of instantaneous frequencies," *IEEE Trans. Inf. Theory*, vol. 38, no. 2, pp. 644–664, Mar. 1992.
[8] P. Guillemain and R. Kronland-Martinet, "Characterization of acoustic signals through continuous linear time-frequency representations," *Proc. IEEE*, vol. 84, no. 4, pp. 561–585, Apr. 1996.
[9] K. Kodera, R. Gendrin, and C. Villedary, "Analysis of time-varying signals with small BT values," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 26, no. 1, pp. 64–76, Feb. 1978.
[10] E. Chassande-Mottin, F. Auger, and P. Flandrin, , L. Debnath, Ed. , "Time-frequency time-scale reassignment," in *Wavelets and Signal Processing*. Boston, MA, USA: Birkhaüser, 2003.
[11] D. Friedman, "Instantaneous-frequency distribution vs. time: An interpretation of the phase structure of speech," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP'85)*, 1985, vol. 10, pp. 1121–1124.
[12] V. Gibiat, F. Wu, P. Perio, and S. Chaintreuil, "Analyse spectrale differentielle (A.S.D.)," *Comptes Rendus de l'Académie des Sciences Paris*, vol. 294, pp. 633–636, 1982, ser II.
[13] S. Maes, "The synchrosqueezed representation yields a new reading of the wavelet transform," in *Proc. SPIE OE/Aerosp. Sens. Dual Use Photon.*, Orlando, FL, USA, Apr. 1995, vol. 2491, pp. 532–559.
[14] R. Carmona, W. Hwang, and B. Torresani, "Multiridge detection and time-frequency reconstruction," *IEEE Trans. Signal Process.*, vol. 47, no. 2, pp. 480–492, Feb. 1999.
[15] R. Lyon, A. Katsiamis, and E. Drakakis, "History and future of auditory filter models," in *Proc. IEEE Int. Symp. Circuits Syst. (ISCAS)*, May 2010, pp. 3809–3812.
[16] T. Irino and R. D. Patterson, "A time-domain, level-dependent auditory filter: The gammachirp," *J. Acoust. Soc. Amer.*, vol. 101, no. 1, pp. 412–419, 1997.
[17] B. C. Moore and B. R. Glasberg, "Formulae describing frequency selectivity as a function of frequency and level, and their use in calculating excitation patterns," *Hearing Res.*, vol. 28, no. 2–3, pp. 209–225, 1987.
[18] B. Moore, *An Introduction to the Psychology of Hearing*. New York, NY, USA: Academic, 2003.
[19] B. Mulgrew, "Detecting stationary phase points in the time-frequency plane," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2013, pp. 5353–5357.
[20] F. Auger and P. Flandrin, "Improving the readability of time-frequency and time-scale representations by the reassignment method," *IEEE Trans. Signal Process.*, vol. 43, no. 5, pp. 1068–1089, May 1995.
[21] S. Kay, *Fundamentals of Statistical Signal Processing*, ser. Signal Processing. Englewood Cliffs, NJ, USA: Prentice-Hall, 1998, vol. II, Detection Theory.
[22] S. Strahl and A. Mertins, "Analysis and design of gammatone signal models," *J. Acoust. Soc. Amer.*, vol. 126, no. 5, pp. 2379–2389, 2009.
[23] CSTR US KED Timit database Apr. 2013 [Online]. Available: http://festvox.org/dbs/dbs_kdt.html
[24] R. Baxley, B. Walkenhorst, and G. Acosta-Marum, "Complex Gaussian ratio distribution with applications for error rate calculation in fading channels with imperfect CSI," in *Proc. IEEE Global Telecommun. Conf. (GLOBECOM 2010)*, Dec. 2010, pp. 1–5.
[25] E. Chassande-Mottin, P. Flandrin, and F. Auger, "On the statistics of spectrogram reassignment vectors," *Multidimens. Syst. Signal Process.*, vol. 9, no. 4, pp. 355–362, Oct. 1998.
[26] T. J. Gardner and M. O. Magnasco, "Sparse time-frequency representations," *Proc. Nat. Acad. Sci. (PNAS)*, vol. 103, no. 16, pp. 6094–6099, Apr. 2006.
[27] M. Zivanovic, "Detection of non-stationary sinusoids by using joint frequency reassignment and null-to-null bandwidth," *Digit. Signal Process.*, vol. 21, no. 1, pp. 77–86, 2011.

Bernard Mulgrew (M'88–SM'07–F'12) received the B.Sc. degree in 1979 from Queen's University Belfast.

After graduation, he worked for 4 years as a Development Engineer in the Radar Systems Department at Ferranti, Edinburgh. From 1983 to 1986, he was a Research Associate with the Department of Electronics and Electrical Engineering at the University of Edinburgh, studying the performance and design of adaptive filter algorithms. He received the Ph.D. degree and was appointed to a lectureship in 1987. He currently holds the Selex ES/Royal Academy of Engineering Research Chair in Signal Processing. His research interests are in adaptive signal processing and estimation theory and in their application to radar and audio systems. He is a coauthor of three books on signal processing and more than 80 journal papers.

Dr. Mulgrew is a Fellow of the Royal Academy of Engineering, the Royal Society of Edinburgh, and the IET.